Confirmatory Factor Analysis in Stata

Karl Ho

Structural Equation Modeling

Structural Equation Modeling (SEM) is a comprehensive modeling framework or approach that encompasses factor analysis, path analysis and regressions. It allows testing of unobserved (latent) concepts (construct) and measurement errors in a structural context. One of the features of SEM is to test models using path diagrams, allowing researchers to conceptualize variables (latent or observed) and iterate model building process.

Confirmatory Factor Analysis

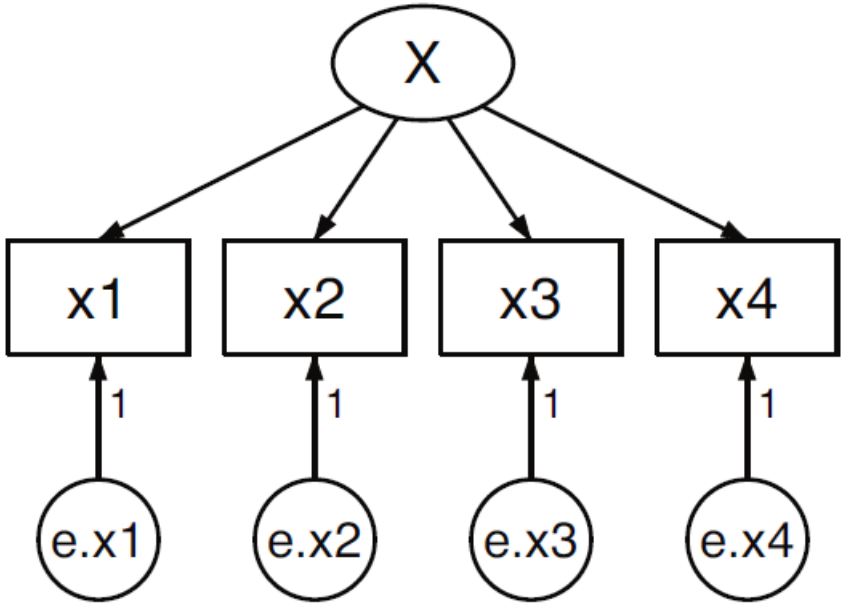
Confirmatory Factor Analysis (CFA) is a theory-driven factor analysis. Unlike Exploratory Factor Analysis (EFA), CFA verifies the factor structure taking into account the measurement errors, theory guided factor loadings and the iterative process of building a "measurement" model.

Why SEM and CFA?

1. **EFA atheoretically provides the number of factors based on different rotation methods and variable communality/uniqueness. It regards no concept of measurement error.**
2. **In the SEM framework, researchers are concerned not only about the number of factors, but also the covariance structure and what we make of the structure. This approach investigates the measurement errors, latent constructs and their structural relations.**

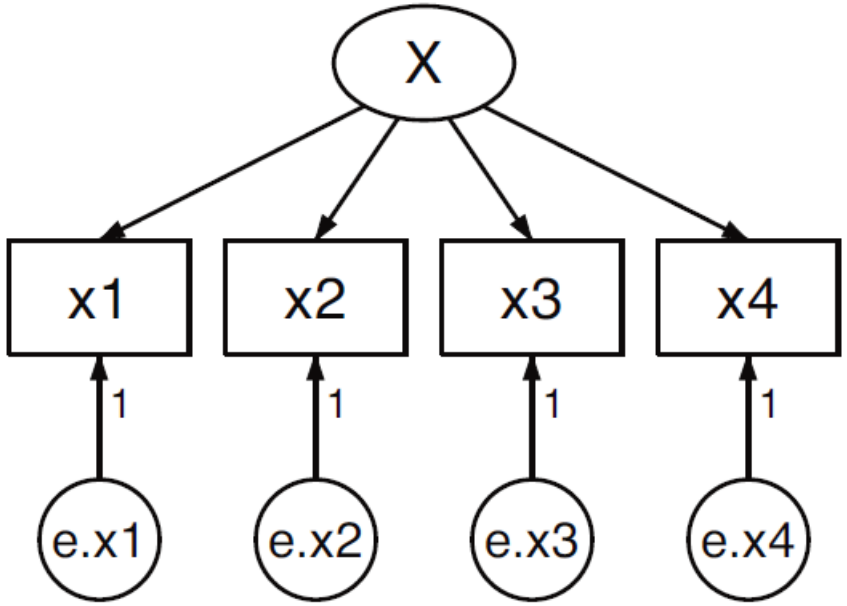
Brief background

Path analysis method dated back to 1920s when geneticist Sewell Wright used boxes, circles and arrows to visualize concepts and their relations.



Source: Stata Structural Equation Modeling Reference Manual, Release 14

Brief background



$$x_1 = \alpha_1 + \beta_1 X + e.x_1$$

$$x_2 = \alpha_2 + \beta_2 X + e.x_2$$

$$x_3 = \alpha_3 + \beta_3 X + e.x_3$$

$$x_4 = \alpha_4 + \beta_4 X + e.x_4$$

Source: Stata Structural Equation Modeling Reference Manual, Release 14

Brief background

Karl Jorëskog introduced methods in analyzing the covariance structure, then the program called LISREL (LInear Structural Relations).

Some terminology:

Latent – unobserved

Construct – observed or unobserved factor

Indicators – observed variables in measurement and structural models

SEM in Stata

Estimating SEM models becomes simplified:

1. Draw a path diagram based on conceptual model
2. Determine indicators, exogenous and endogenous variables
3. Determine path directions (covariance, single direction path)
4. Run estimation

SEM in Stata - Syntax

Stata command:

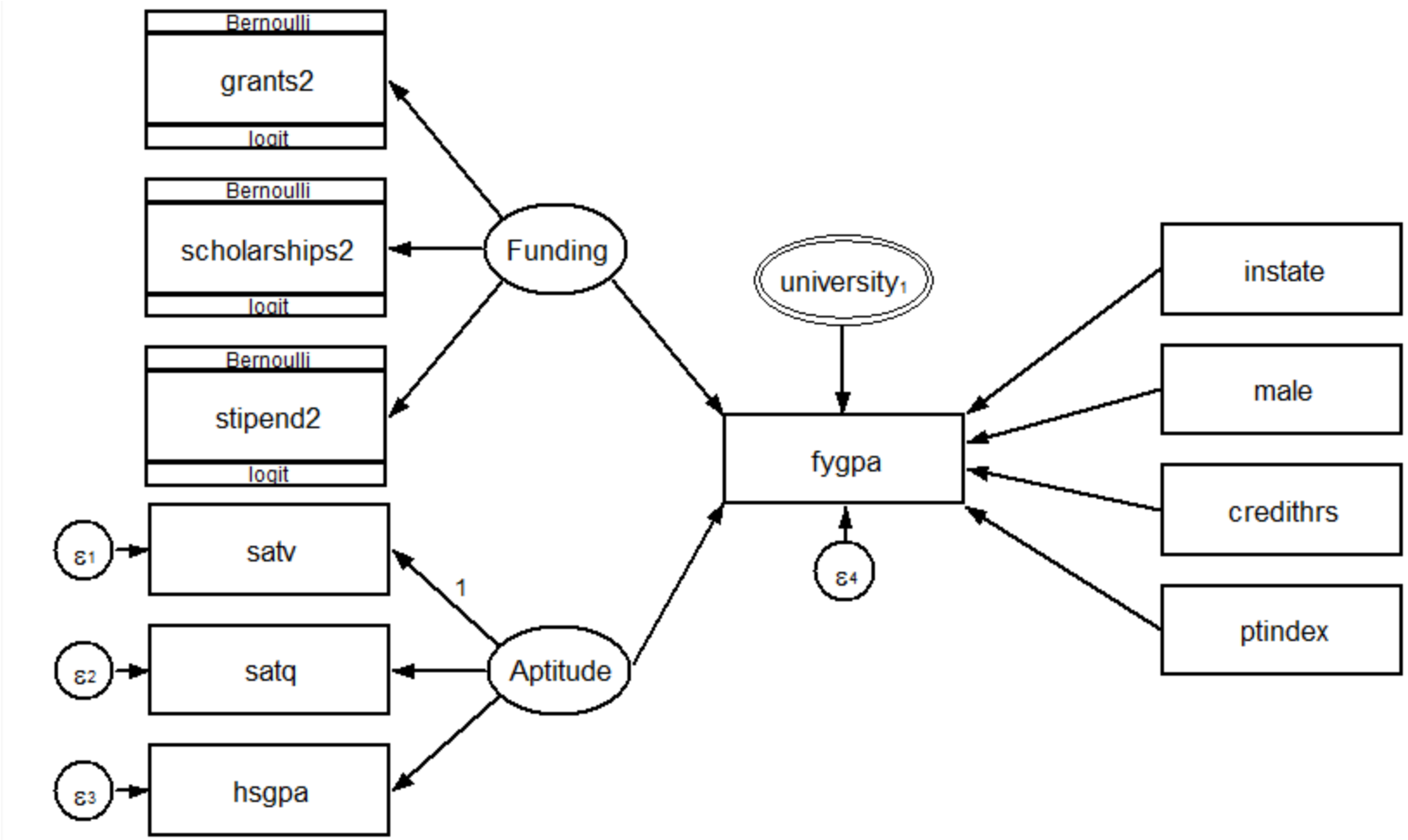
e.g.

sem (EFFICACY-> x1 x2 x3)

sem (x1 x2 x3 <- EFFICACY)

sem (EFFICACY-> x1 x2 x3), cov (e.x1*e.x2) cov (e.x1*e.x3)

SEM structural model



Huber, Chuck. 2014. *Introduction to Structural Equation Modeling Using Stata.* Prepared for presentation at California Association for Institutional Research November 19, 2014

Example: Ho and Tan 2012

A Structural Analysis of Social Network and Political Capital in Democratic Taiwan

Taiwanese voters demonstrate high level of participation in voting but low rate of members in civil organizations (Ho and Tan 2006; Kim 2005). Unlike citizens in Western democracies, Taiwanese voters are less likely to accumulate their social and political capital among members of formal social networks. It does not infer that citizens in the young democracy are not keen in political mobilization via social and political networks. In this study, we investigate more closely how political capital is accumulated within and beyond social and political networks. To continue an effort in building a conceptual framework of the political capital theory, we examine how voters form and convey their political opinions through social and political networks and turn to political actions including cognitively seeking political information and actively partaking electoral participatory activities. In a broader sense, we intend to build a theory that explains political *transactions* at different levels of a civil, democratic society. By political capital, we refer to in a general sense the individual's ability and capacity to engage in political actions and decision making. (Sørensen and Torfing 2003; see also Fuchs et al. 2001 and Booth and Richards 1998). This concept entails three components: 1. the level of access the general citizens have to decision-making processes (endowment); 2. their capability to make a difference in these processes (empowerment); and 3. their perception of themselves as political actors (engagement). Using data from the 2012 Taiwan Election and Democracy Study, we construct a structural model of political capital and build on a capitalist theory to explain social connections and voting decision process. We introduce a new theoretical framework in explaining acquisition of political information and political actions.

Political Capital – Endowment

1. **Political discussions (poltalk)**

   Informed by the studies on political discussions and political participation (McClurg 2006; Klofstad, McClurg and Rolfe 2009), we use the question asking how often the respondent talks with other people about politics to measure the political network concept. This four-point scale variable ranges from a zero (never) to four (often).

Political Capital – Endowment

2. Political Network Size (polnetsize)

The second question is “who do you commonly talk about politics with” and the respondent can give multiple responses ranging from family members to members of an association. We count the number of these “network others” and generate the political network size variable.

Political Capital – Endowment

3. Network Heterogeneity (heteronet)

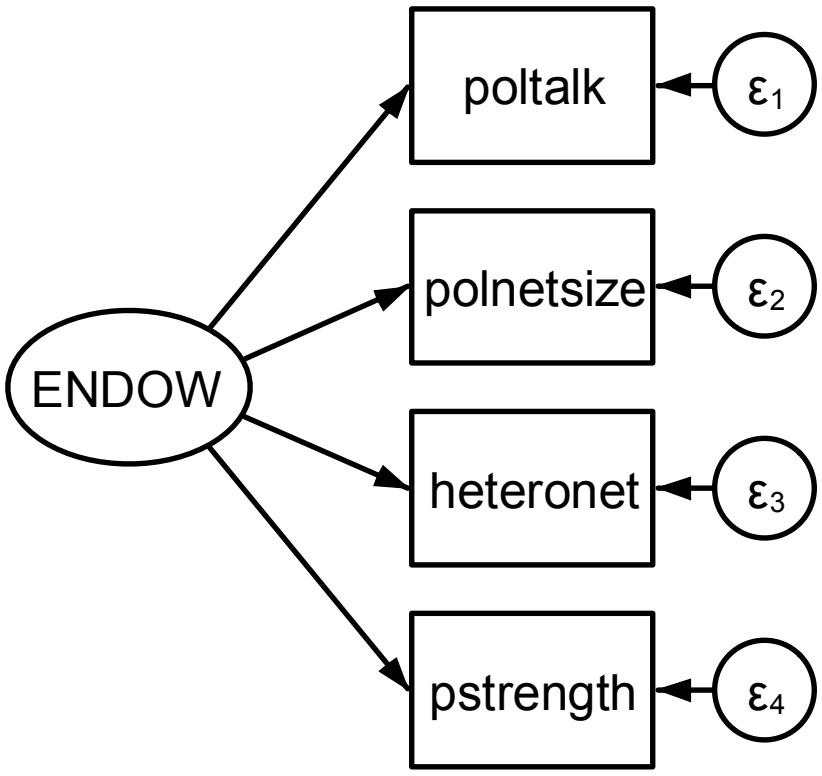
Empirical studies on social political networks suggest cross-cutting network exchanges or heterogeneous networks could contribute to building of tolerance in a civic society but may not necessarily encourage political participations (Mutz 2002). Such network heterogeneity may also conducive to lower turnout (Liu and Chiu 2011). While this concept of accommodating disagreement with political discussants may help build the endowment component by possibly enlarging political network, it can be orthogonal to the party support component (engagement). We measure this variable by follow-up question to the political discussion questions asking if the people who the respondent commonly discusses politics or elections with support the same party. A five-point scale is developed ranging from all support same party (least heterogeneous network) to none (most heterogeneous network).

Political Capital – Endowment

4. Associational Activeness (pstrength)

This construct taps into how active the respondent is participating in activities of associations. This constitutes the social capital concept of formal social networks. We use the strength of party support as a proxy of the social capital component and gauge how this variable is associated with other political capital components.

Political Capital – Endowment



Political Capital – Endowment
EFA

```
. factor poltalk polnetsize heteronet pstrength
(obs=1,826)

Factor analysis/correlation                      Number of obs    =      1,826
    Method: principal factors                    Retained factors =          2
    Rotation: (unrotated)                        Number of params =          6
```

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	1.75243	1.65311	1.1505	1.1505
Factor2	0.09932	0.23464	0.0652	1.2157
Factor3	-0.13532	0.05794	-0.0888	1.1269
Factor4	-0.19326	.	-0.1269	1.0000

LR test: independent vs. saturated: chi2(6) = 2176.14 Prob>chi2 = 0.0000

Factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Uniqueness
poltalk	0.8093	0.0521	0.3423
polnetsize	0.7648	0.0063	0.4150
heteronet	0.6803	-0.1570	0.5126
pstrength	0.2229	0.2682	0.8784

Political Capital – Endowment
EFA

Factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Uniqueness
poltalk	0.8093	0.0521	0.3423
polnetsize	0.7648	0.0063	0.4150
heteronet	0.6803	-0.1570	0.5126
pstrength	0.2229	0.2682	0.8784

Political Capital – Endowment

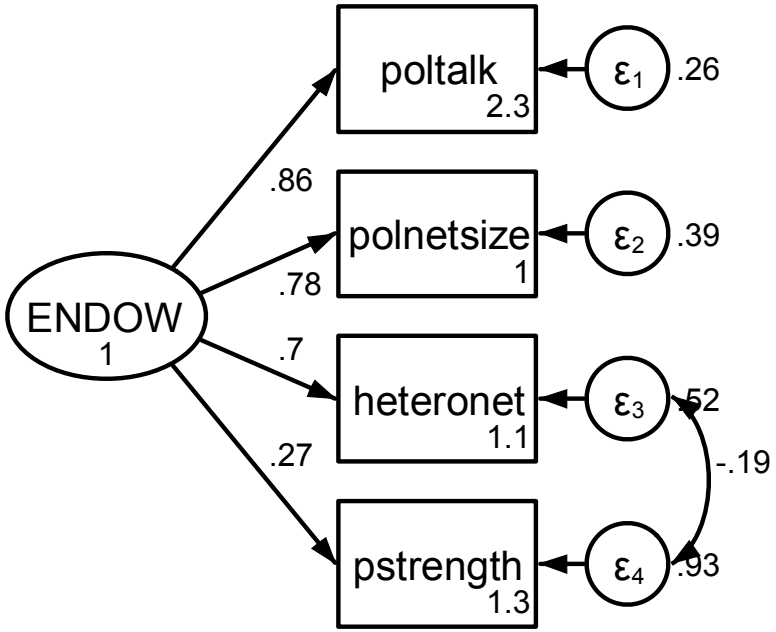
Note:

Identification of model

Model fit

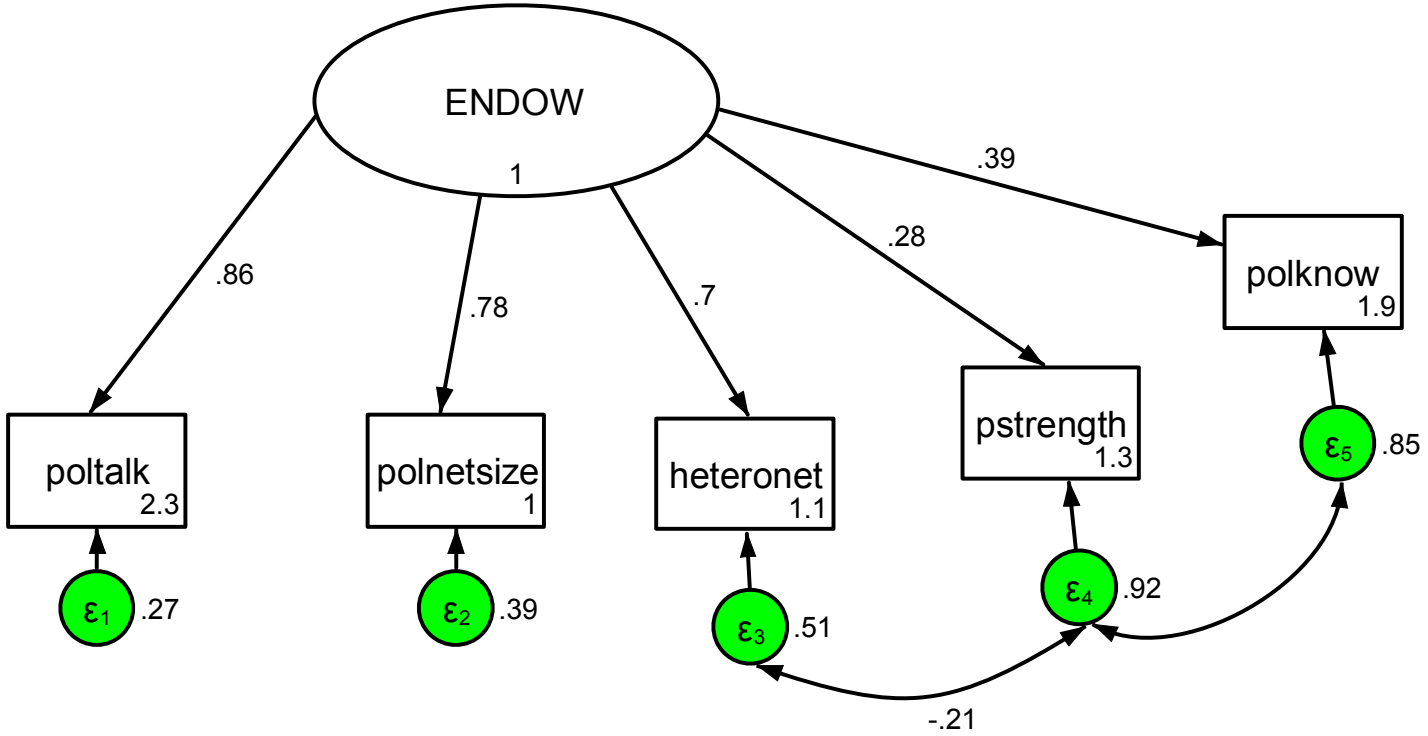
- Most common index of fit is χ^2 goodness-of-fit index
- Really a “badness-of-fit” index. Small numbers mean better fit (0 would be best of all)
- Small p values (e.g. < .05) are bad
- To improve model fit:
  - Reviewing the structure
  - Reviewing the modification indices, any more path to estimate to get better fit

Political Capital – Endowment



LR test of model vs. saturated: chi2(1) = 5.49, Prob > chi2 = 0.0191

Political Capital – Endowment



LR test of model vs. saturated: chi2(3)=8.60, Prob > chi2 = 0.0351 CFI=0.989 RMSEA=0.062

Democracy example

CFA:

One factor model

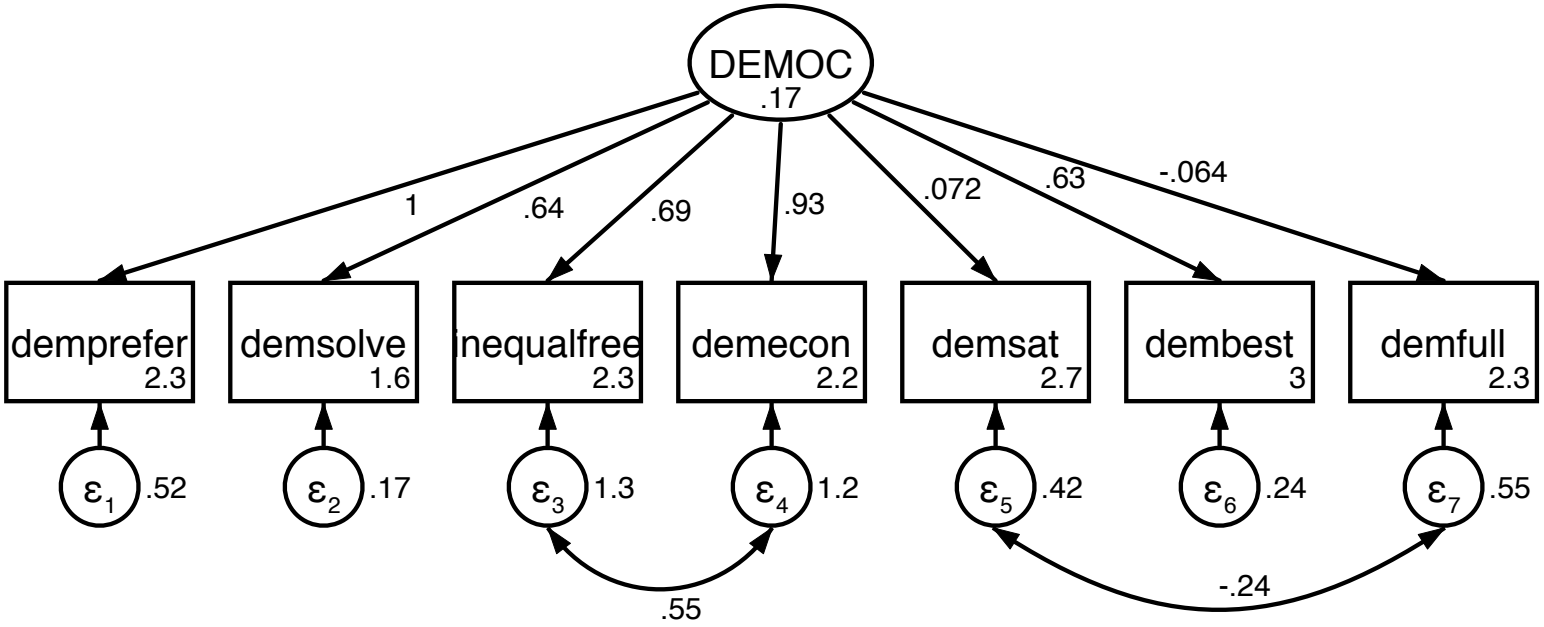
Two factor model

Structural model

Democracy example

CFA:

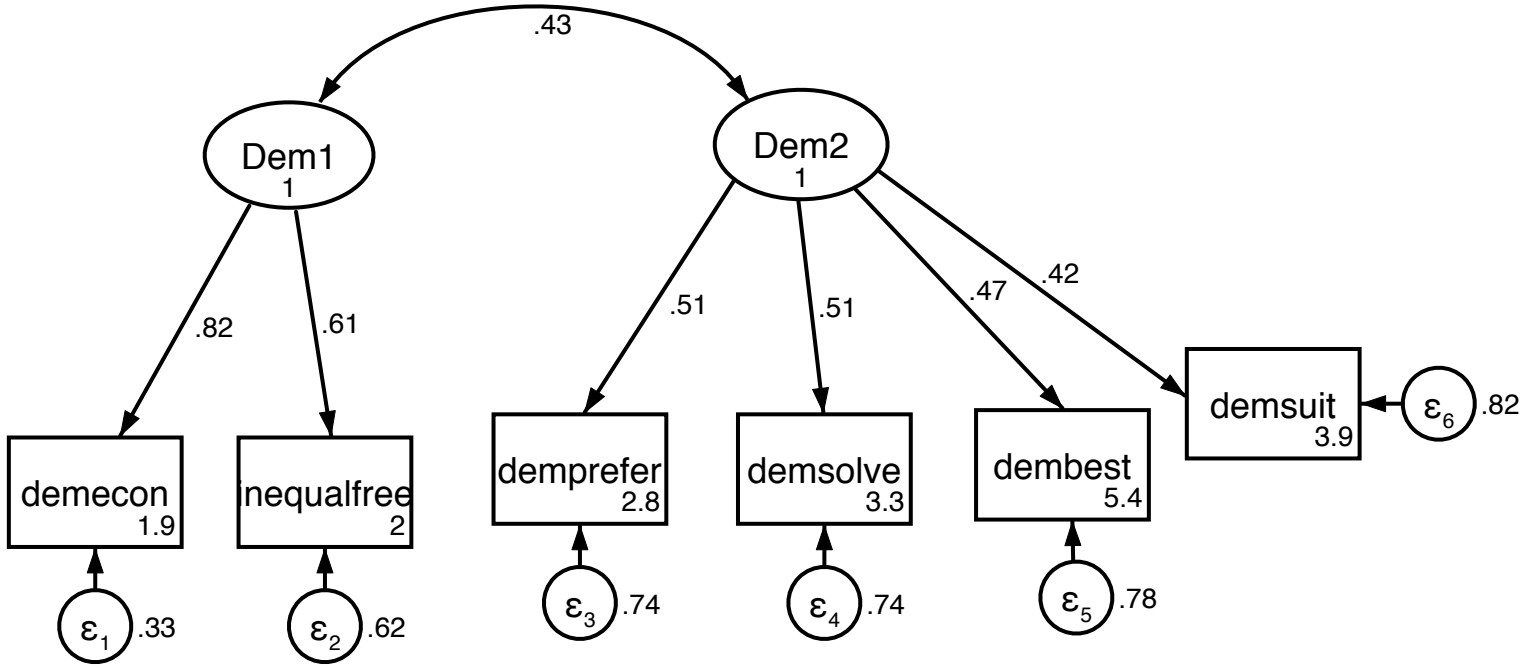
One factor model



Democracy example

CFA:

Two factor model



Democracy example

Structural model: